

實現一個新型高效率的音樂合成器--應用 HHT 時頻分析法

趙春棠

南台科技大學 電機系
tang@mail.stut.edu.tw

蕭逸豪

南台科技大學 電機系
M9720214@webmail.stut.edu.tw

摘要

電腦音樂合成在現代音樂創作中，佔有極重要的地位。一個好的音樂合成法，能讓音樂創作者在最短的時間內，快速且精準地合成所要產生的樂音。本論文應用 HHT 時頻分析法，實現一個新型高效率的音樂合成器。鋼琴、長笛、以及小提琴等樂器之樂音訊號，作為 HHT 時頻分析法之輸入，藉以協助找出新型有效率的樂音合成法。實驗發現只要能針對 HHT 時頻分析法所得之 IMF 及其瞬時頻率，做有效的合成，即可合成出接近原樂器音色的品質，本論文之研究成果，將可進一步應用於音樂壓縮等技術。

關鍵詞：音樂合成、HHT、經驗模態分解、內建模態函數。

1. 前言

一般電子音樂合成法中，最常被應用的有波表法 (Wave Table)，以及調頻法 (FM)。然而這兩種方法，已經不能滿足現今越來越嚴苛的品質要求。電腦音樂合成在現代音樂創作中，佔有極重要的地位。一個好的音樂合成法，能讓音樂創作者在最短的時間內，快速且精準地合成所要產生的樂音。近年來的音樂合成研究趨勢乃是試圖以實體模型法 (Model-Based Methods) [1]，如數位波導濾波器 DWF (Digital Waveguide Filter) 來模擬實體樂器的物理特性，如圖 1 可以作為一般樂器的模型，其中激發單元 (Exciter) 是非線性的部分，負責產生振盪訊號源；而共振單元 (Resonator) 則是屬於線性的部分，負責將振盪訊號傳導出去。如此一來，就有許多研究利用類神經網路及最佳化等技術，去針對不同的樂器，學習出以下的架構參數。圖 2 則為 IIR(Infinite Impulse Response)合成架構中的重複性網路型式[2]，其中亦利用了類神經網路作學習，利用 RNN(recurrent NN)設計預估濾波器 (Prediction Filter)。對於這種設計一良好的模式為主的合成法，需要適合的激勵輸入方能產生一逼真的合成音。而事實上，要精確的選擇一個樂器的模型及參數是困難重重的，且不易控制，尤其在了解其非線性驅動的物理意義之下，更難合成出樂器的原訊號。而且利用類神經網路學習，其費時及使用大量參數，是可想而知的！

本論文希望藉由對希爾伯-黃轉換 (Hibert-Hung Transform, HHT)時頻分析法的了解，能研發一套更有效率的合成方法。

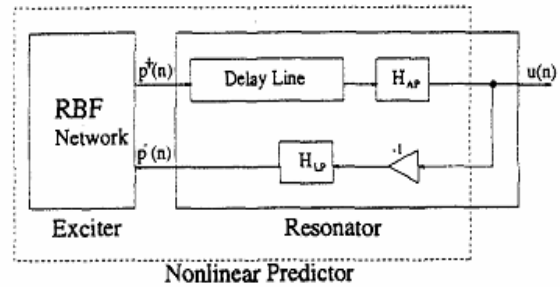


圖 1. 樂器的非線性預估模型

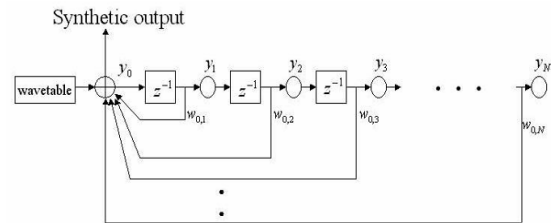


圖 2. IIR 合成架構[2]

2. 實驗方法

大部分信號皆以波的形式來表示，音樂信號亦不例外。說到波形，不免要談到振幅、頻率等等。而像音樂訊號的波形，我們可以在不同的時間區段上找到不同的頻率週期，造成這種現象的主要原因，是因為音樂訊號本身是由很多非線性的訊號所組成。這種隨時間變化的訊號藉由 HHT 經驗模態分解技術，隨時變動音高的音樂訊號，將分解成許多內建模態函數，希望藉由這些不同樂器的聲紋特徵，來實現一個新型高效率的音樂合成器。

2.1 實驗流程

圖 3 所示，為本論文之實驗流程。

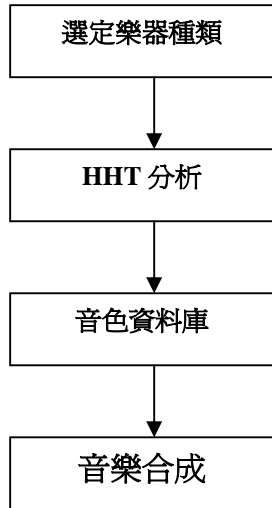


圖 3. 流程圖

2.2 選定樂器種類

由於樂器種類繁多，而為了驗證本合成方法的有效性，我們先利用大眾化的樂器如鋼琴，小提琴等，作為研究的方向。圖 4、5、6 分別為鋼琴 A4、小提琴 A4、長笛 A4 彈奏一秒的時間波形[3][4]。待研究成果顯著後，再推廣至所有其他樂器。

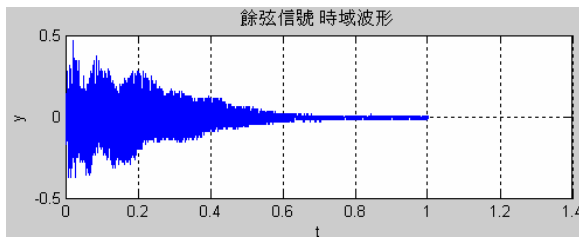


圖 4. 鋼琴演奏 A4 音高之波形

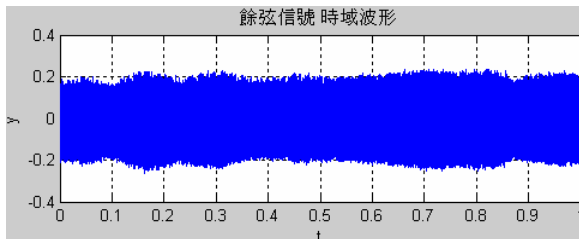


圖 5. 小提琴演奏 A4 音高之波形

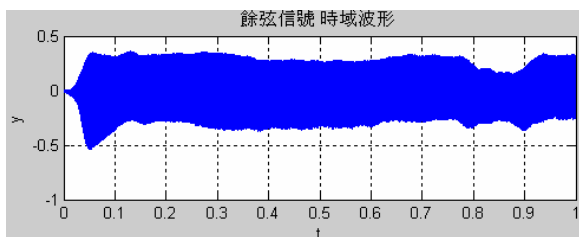


圖 6. 長笛演奏 A4 音高之波形

2.3 樂器音色的 HHT 分析，重要參數的萃取

這部分在過去的作法，目的在取得基頻、音量、及雜音等音量參數。而對於特別樂器的特殊演奏技巧，例如古典吉他與二胡都具有的揉音、顫音、或滑音等，困難度可想而知[5]。本論文利用 HHT 分析，則希望能獲得更有效的參數，以取代過去傳統的作法。例如樂音頻率大多呈現倍數存在的特性，光是尋找基頻，就非易事。這些問題都希望能利用 HHT 時頻分析技術，可以得到更有效率的分析技術。

2.3.1 希爾伯特-黃轉換(HHT)

HHT 的是一種非線性分析工具，它是適用於非穩態 AM/FM 的數據，例如準週期的音樂信號。尤其，HHT 的局部分析能力讓我們能夠分析隨時間變化的信號，在更短時間的視窗，所以 HHT 在樂音的音高判斷分析，提供了很好的工具[6]。

HHT 計算包含兩個部份：驗模態分解法 (Empirical Mode Decomposition, EMD) 和 Hilbert 轉換，EMD 又稱為 Sifting，可作為預先處理希爾伯特變換的步驟，它採用了非線性程序，將一訊號分解成數個 IMF (Intrinsic Mode Functions)，震動模態是一個函數需滿足以下兩個條件：

- (1) 零點和極值的數量相差在一個以下。
- (2) 任一處的上包絡線 (由局部極大值定義) 和下包絡線 (由局部極小值定義) 的平均值為零 (即對稱)。IMF 可視為是廣義的傅立葉轉換，隨時間變化的振幅和瞬時頻率大大提升了計算效率。

2.3.2 EMD

HHT 在非穩態 AM-FM 等信號是一個強有力的分析工具[7]，輸入信號拆解成簡單的 AM-FM 窄頻信號以至於 AM-FM 信號利用希爾伯特變換不會干涉其它 AM-FM 信號。

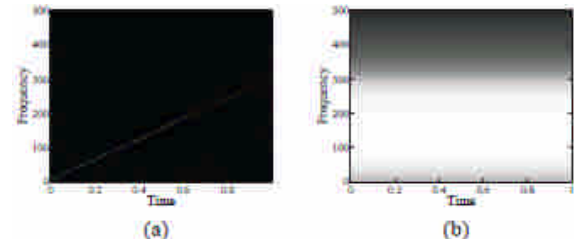


圖 7. 線性調頻信號的時頻 $\omega(t) = 300 \times 2\pi t$ 獲得從 (a)HHT 和 (b) 傅立葉變換。

比較 HHT 頻譜和傅立葉轉換隨時間針對一小段的線性調頻信號的成分。觀察發現，HHT 顯示比傅立葉轉換有更好的解析度，其中白色代表著比較高能量。尤其 HHT 的局部分析能力在更短時間的視

窗讓我們能夠分析隨時間變化的信號。圖7說明在樂音瞬時音高之的分析上，HHT提供了很好的工具。

EMD的本意，是一個可以將一訊號分解成數個IMF和一個剩餘訊號的數學工具。IMF可視為是廣義的傅立葉轉換，隨時間變化的振幅和瞬時頻率大大提升了計算效率，並讓我們可以得到訊號中非線性和非穩態的特性。計算IMF時，振幅資訊和頻率資訊將被分離開來，因此克服了傅立葉轉換中固定振幅和固定頻率的限制，得到則是隨時間變化的振幅和頻率關係。EMD可依據訊號特性將一原訊號分解為多個子訊號。而 Sifting 是做EMD計算時，從訊號中提取IMF，Sifting的運算流程如下[8]：

1. 取得一段要處理的訊號 $s(t)$ 如圖8。找出其局部極大值所在點，以cubic spline方法將所有局部極大值相連，可得此訊號的上包絡線如圖9藍色線所示。
2. 同樣的找出此訊號的局部極小值以cubic spline方法相連，可得訊號的下包絡線如圖9紅色線所示。
3. 計算上包絡線和下包絡線的平均值，可得一平均線，令此平均線為 $m(t)$ 如圖10所示。
4. 將原訊號和平均線 m 相減，令相減後所得的訊號為

$$h(t) = s(t) - m(t) \quad (1)$$

上述由 $s(t)$ 減去上下包絡線的平均線，如式(1)，求得 $h(t)$ 的過程即為sifting如圖11。式(2)是一個終止準則來根據停止sifting，就是當成分 $C_n(t)$ 或殘餘訊號 $r_n(t)$ 訊號過零點個數和極值個數之差小於1為止；或殘餘訊號成為單調函數不再可以提取IMF，sifting動作就結束。

$$s(t) = \sum_{i=0}^n c_i(t) + r_n(t) \quad (2)$$

圖12、13、14分別為鋼琴A3、小提琴A4及長笛A4作EMD的結果，影響樂器的音主要是在IMF1到IMF4，所以我們就利用原音分解過後此4個音來萃取的重要參數以達到完成合成音樂的依據。

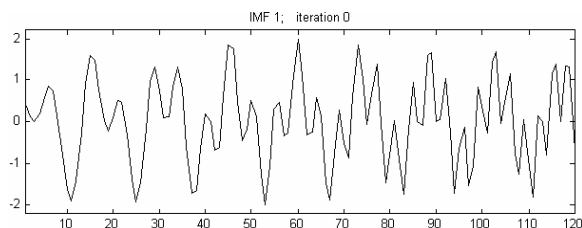


圖8. 原始振動訊號

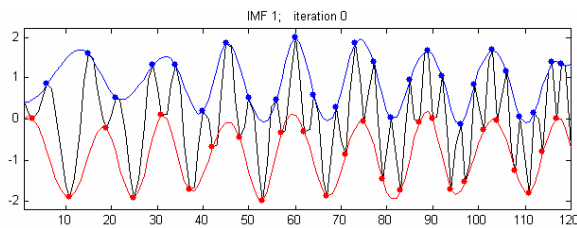


圖9. 極值上包絡線與下包絡線

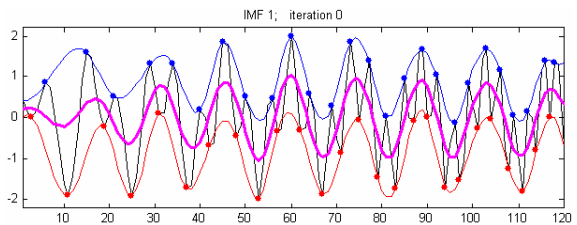


圖10. 均值包絡線

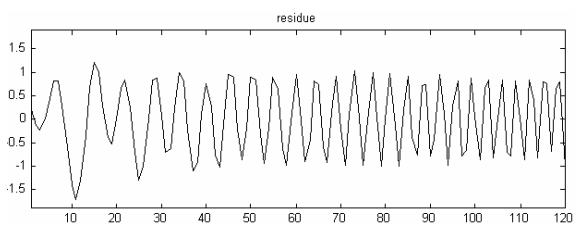


圖11. 第一次轉移過程所得分量 h_1

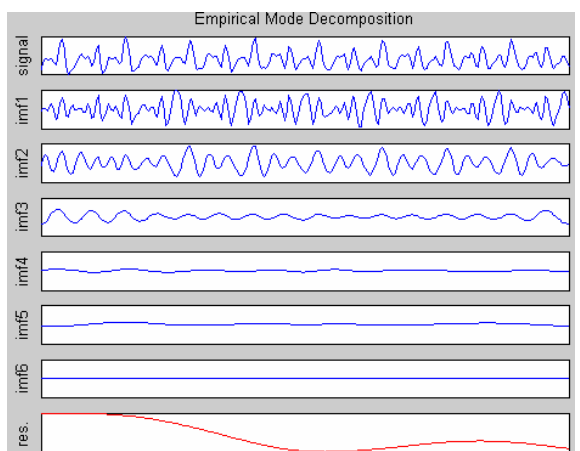


圖12. 鋼琴A4之EMD分析

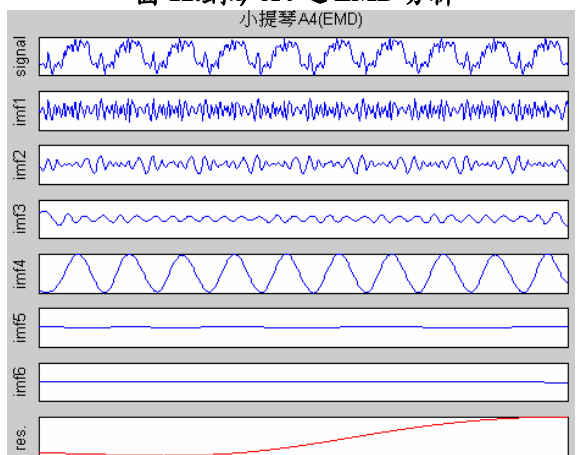


圖13. 小提琴A4之EMD分析

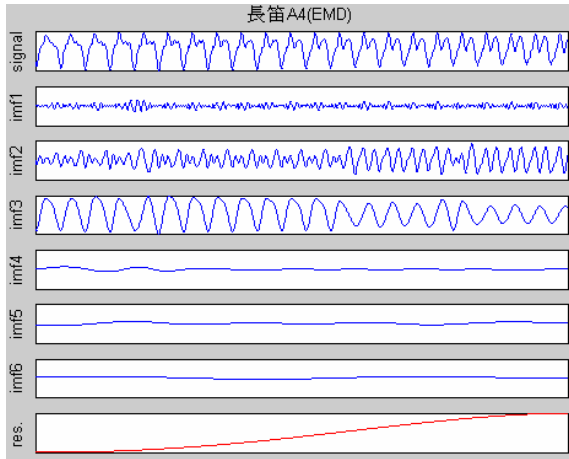


圖 14. 長笛 A4 之 EMD 分析

2.3.3 希爾伯特頻譜分析

在取得固有模式函式成分，我們應用 HHT 變換對每個 IMF 成分，並計算瞬時頻率。之後執行了 HHT 變換在每個 IMF 的成分，原始數據可以以實部 \Re 表示為式(3)

$$s(t) = \Re \left\{ \sum_{j=1}^n a_j(t) \exp[i \int \omega_j(t) dt] \right\} \quad (3)$$

在這 a_j 是瞬時幅度， ω_j 是瞬時頻率而 IMF 有個殘留訊號 $r_n(t)$ 可是在這裡被忽視，為什麼殘留訊號可以被忽視，因為根據了篩選停止準則，以致不再提取 IMF，這裡的殘餘在信號分析是無用的，殘餘是沒有物理意義。因此，殘餘訊號是可以忽略的。在 EMD 是基於信號是由任何許多不同 IMF 所構成。然而，希爾伯特變換可用於識別瞬時頻率。

式(3)使代表振幅和瞬時頻率作為時間的函數，可以畫成在三維圖上，如圖 15 為鋼琴 A4 的三維圖，其中振幅可以顯示頻率在時間平面剖面圖。這種振幅的頻率時間分佈的被指為希爾伯特振幅頻譜 $H(\omega, t)$ 。簡單地來說就是希爾伯特頻譜。

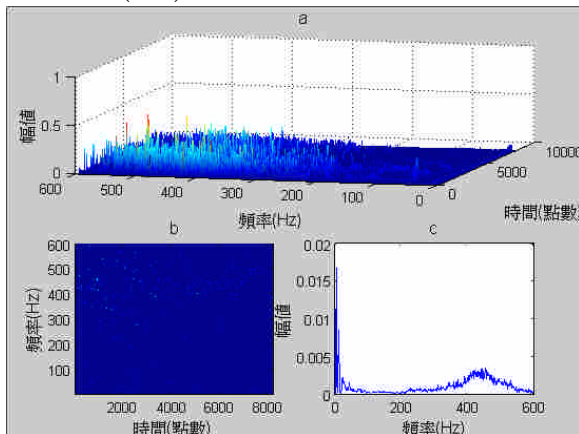


圖 15. 鋼琴 A4 三維圖

定義了 Hilbert 頻譜，我們也可以定義邊際頻譜

$h(\omega)$ 如式(4):

$$h(\omega) = \int_0^T H(\omega, t) dt \quad (4)$$

邊際頻譜給予了全部振幅的分量（或能源）貢獻於每個頻率值[9]。圖 16 也代表了鋼琴 A4 振幅累計在整個數據範圍概率意義。

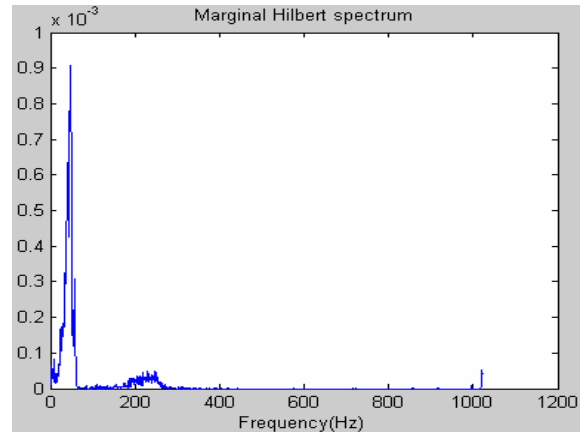


圖 16. 鋼琴 A4 Hilbert 邊際頻譜

訊號經 EMD 得到的各 IMF 它們包含了節奏部分的基礎音，因此我們可以從各 IMF 來得到瞬時頻率，就可以了解基音、節奏音的主音，使達到音樂合成的依據來源。圖 17、18、19 分別為鋼琴 A4、小提琴 A4 及長笛 A4 的瞬時頻率。

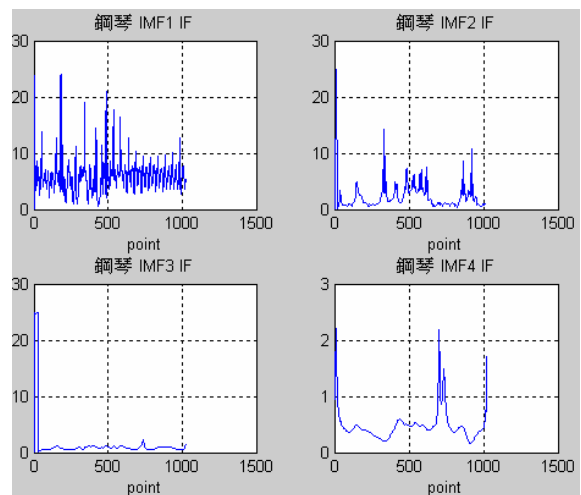


圖 17. 鋼琴 A4 之 imf1_2_3_4_IF

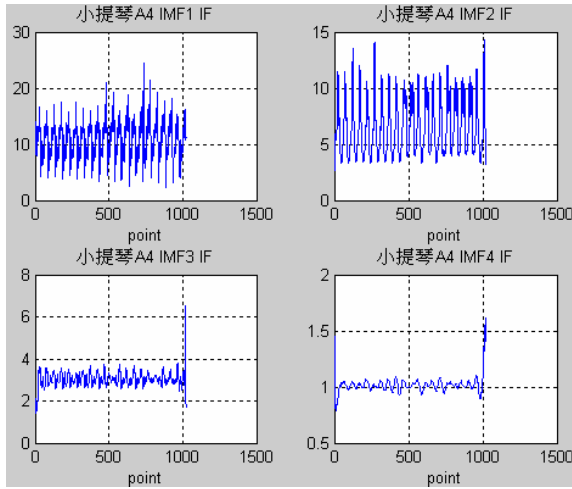


圖 18. 小提琴 A4 之 imf1_2_3_4_IF

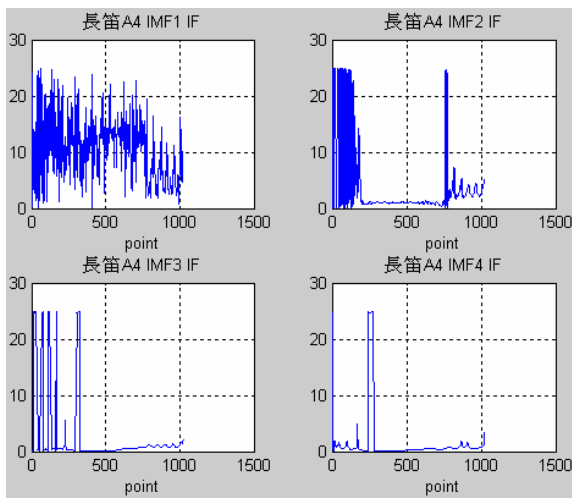


圖 19. 長笛 A4 之 imf1_2_3_4_IF

3. 音色庫的建立

音樂的音調包括一個基本音調，以及一系列的諧波；基本音調即某一固定的振動頻率，每個諧波的頻率為基本頻率的整數倍。每個諧波包含的能量決定於發聲的樂器種類，以此分辨各種樂器。音樂中將頻率稱為音調(Pitch)；聲音的特性決定於包含的諧波成分，稱為音色(timbre)。諧波亦稱為倍音(overtone)或泛音(partial)。音色庫應包含諧音音色庫以及雜音音色庫兩大部分。內容應記錄每個諧音在特定基頻及音量時的強度，以及定音高雜音的頻譜。音色庫能充分代表一個樂器的演奏特色，如果能設計得當，自然能合成出理想的音色；而聲音在 20Hz--60Hz 這一段提升能給音樂強有力的感覺，給人很響的感覺，如雷聲。如果提升過高，則會混濁不清，造成清晰度不佳，特別是低頻回應差和低頻過重的音響設備。

4. 樂音的合成

4.1 樣本播放合成

這方法於現今一般音效卡所採取的合成方式，利用儲存在硬碟裡的音色庫，可讓使用者更換喜愛的樂器音色。取得樣本後再經過調整音高、音長來得到合成的訊號。而合成的步驟如下：

1. 從取樣波形中擷取一個或多個循環點。
2. 依照音長定下所需的起點與停止點。
3. 用傳統的波形封包來調整波形振幅。
4. 加入其他合成元件，譬如以濾波器修飾音色。
5. 將所有合成係數紀錄到記憶體中。

樣本撥放合成方法很適合應用到本論文中且樣本播放合成的特色就是可以有如同原音的效果，但其原理就是採先錄音後撥放，需儲存大量高解析度的取樣波形，將造成資料量龐大，本論文即是希望藉由HHT分析法，能研發出新且有效率的合成法。

4.2 傅立葉合成

數學家傅立葉證明了幾乎任何連續函數可以產生無限的正弦和餘弦波的總和，其結果深遠的影響聲音的合成和重現。純正弦波可以轉換成聲音的喇叭，並被認為是一個穩定的、清亮且是唯一的音調 pitch。管弦樂器的聲音通常包括基本音和諧波，這可以被視為是一個正弦波疊加的基本頻率 f 和整數倍數的頻率。分解樂器聲音或任何其他週期函數成其構成正弦或餘弦波的過程被稱為傅立葉分析[10]。

就描述就聲波的特性而論，實現正弦波組成的振幅；這組數字告訴你的聲音的諧波含量，有時稱為諧波聲音頻譜。諧波的內容最重要是限定持續音符的品質和音色。從傅立葉分析一旦知道持續的音樂聲的諧波含量，就有能力從一系列純音正確地適當調整振幅和相位分階段加在一起合成。這就是所謂的傅立葉合成。其中一個重要的概念聲音重現由傅立葉分析引起，傅立葉分析是採用高品質的音頻重現系統去重現撞擊的聲音或快速瞬變的聲音。比如長號的持續的聲音可以重現頻率在有限範圍，因為大多數的聲能是在基本 pitch 的第一數諧波。但如果你將要合成鈸的半音起奏，你需要一個廣泛的高頻使產生快速變化。你可以設想把正弦波產生半音的脈衝加起來，或許可以看到您需要大波的振幅與上升在很短時間（高頻率），以重現鈸的高音起奏。傅立葉分析可以概括地說，任何聲音的高音起奏，或高音脈衝，或在快速變化的波形如方波將有很多高頻率的內容。

傅立葉級數分析能夠使週期函數表示為正弦的無窮三角級數和 \cos 條件。我們知道，音樂聲音可以分析量測功率在合聲的分部，對於某一些特殊

的樂器，這種分部是合成聲音的關鍵。給定一個基本頻率 f_1 和功率組合 p_n 與 n 項諧波，合成波形是

$$y(t) = \sum_{n=1}^N p_n \cos 2\pi n f_1 t \quad (5)$$

再利用 HHT 分解得到的各 IMF 能量在一個基頻以及諧波頻率，並將原音利用傅立葉分析得到的波函數(通常是振幅、頻率或相位隨時間變化)可以表示為一個總合的正弦和餘弦函數，如圖 20、21、22 分別為鋼琴 A4、小提琴 A4 及長笛 A4 合成過後所得到波形，可以發現與原音相去不遠，由此利用 HHT 分解原音得到的 IMF 是可以有效的幾近原音重現。

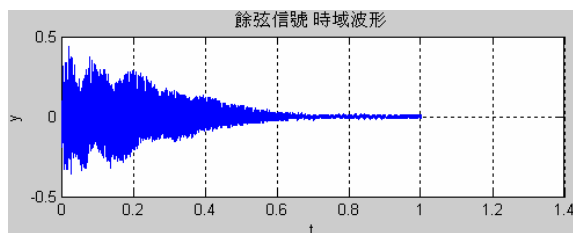


圖 20. 鋼琴 A4 合成

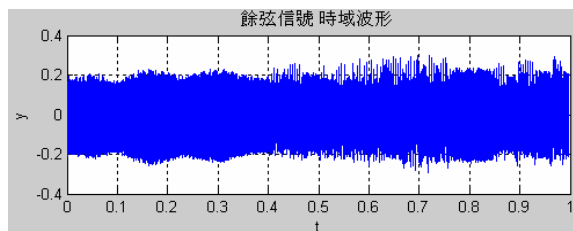


圖 21. 小提琴 A4 合成

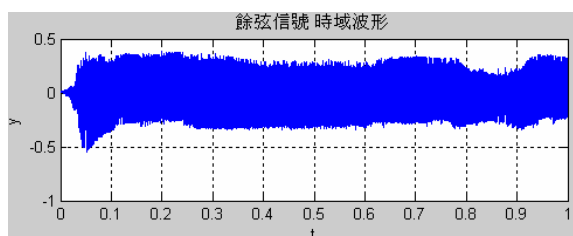


圖 22. 長笛 A4 合成

5. 結論

樂器的每個音都包括一個基本頻率和其諧波。用數學方法分析一個複雜的波形，決定其包含成分的頻率、振幅和相位，並利用 HHT 萃取音樂重要的參數就可以合成原音。而具有瞬時特性的聲音例如鋼琴、鼓或鑼發出的聲音，重現時必須清脆生動。像是鑼突然碰撞一下產生的波形前端非常陡，因為振幅突然增加，包含有大量的諧波。如果這些諧波不能有效地再生，有失真或損失，則音樂會缺乏「魄力」或「衝擊力」。如此，一個不懂

樂器的人不需花多年的功夫鑽研，只需幾分鐘時間就可以彈奏出各種美妙的音樂，如交響樂一般的多樣化！依據音色庫裡查詢不同樂器的聲紋特徵，即可有效率且快速的合成各種樂器聲音，有如實際樂器所彈奏出的音樂，而特殊演奏技巧如揉音、抖音、或滑音也可以完成。希望藉由本論文的實現，可以提升科技人員的人文藝術修養，帶動科技與藝術之結合。

致謝

本論文承蒙行政院國科會計畫補助經費(計畫編號: NSC 98-2221-E-218-030-)，得以順利執行完成，特此致謝！

參考文獻

- [1] 王瑞文，實體模擬技術運用於撥弦樂器、滑音處理與古琴輸入法之人機介面研究，中華大學資訊工程所碩士論文(2001)
- [2] 張瑋倫，頻域音樂合成法-應用於二胡，成功大學資訊工程所碩士論文(2007)
- [3] 張智星，MATLAB 程式設計【入門篇】，鈦思科技出版社
- [4] 梁伯達，洞簫音色之 Hilbert-Huang Transform(HHT)分析，臺灣大學電信工程學研究所論文(2007)
- [5] C. Drioli and D. Rocchesso, "A generalized musical-tone generator with application to sound compression and synthesis," Proceeding of the IEEE International Conference on Acoustics, Speech, and Signal Processing, Vol. 1, pp. 431-434, 1997.
- [6] E. Tsau; N. Cho and C.J. Kuo, "Fundamental Frequency Estimation for Music Signals with Modified Hilbert-Huang Transform". IEEE International Conference on Multimedia and Expo, 2009. ICME 2009, Page(s): 338 - 341
- [7] M. R. Petersen, "Musical Analysis and Synthesis in Matlab" appearing in the MAA's College Mathematics Journal, Vol. 35, No. 5, November 2004, p.396-401
- [8] Huang, "The Empirical Mode Decomposition and the Hilbert Spectrum for Nonlinear and Non-Stationary Time Series Analysis," Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, vol. 454, no. 1971, pp. 903-995, March 1998.
- [9] F. Liu; X. Li, and Z. Liang ;, "Short-wave Aviation Communication Signal Analysis and Aircraft Classification", 2008 International Multi-symposiums on Computer and Computational Sciences, Page(s): 114 - 118
- [10] <http://www.phy.ntnu.edu.tw/java/sound/>